Algoritmo Winowing:

Idea principal: Este algoritmo es usado como parte de un método para hallar las huellas (fingerprint) de un documento dado.

En general un método que sigue el esquema del fingerprint trata de encontrar fragmentos significativos del un texto dado, que permitan realizar una comparación con los fragmentos obtenidos y encontrar coincidencias.

El algoritmo winoming general es:

* Dado un texto, obtener las palabras.
* Dadas las palabras se deben obtener particiones, donde una partición se define como K palabras consecutivas, finalmente, se convierte a cada partición en un entero.
* Finalmente para la lista de particiones, se utiliza una ventana de tamaño w de cada ventana cogemos al menor entero, en caso hayan 2 o más, coger el de más a la derecha.

Ejm w = 4

De la secuecia: 45 25 69 82 41 35 14

Las ventanas serian

45 25 69 82

25 69 82 41

69 82 41 35

82 41 35 14

Los seleccionados serian: 25 25 35 14

Nota: Generalmente para ahorrar espacio si 2 ventanas consecutivas dan el mismo “representante” se considera solo el primero. En el ejemplo, quedaría : 25 35 14.

Entonces para comparar 2 documentos , se obtienen esos valores representativos de cada uno de ellos, y se compara la cantidad de valores en común para determinar si existe plagio o no.